AD NUMBER

ADB225895

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies only; Specific Authority; 11 Jul 97. Other requests shall be referred to Commander, U.S. Army Medical Research and Materiel Command, Attn: MCMR-RMI-S, For Detrick, Frederick, MD 21702-5012.

AUTHORITY

USAMRMC ltr, 4 Dec 2002

AD_____

CONTRACT NUMBER DAMD17-97-C-7018

TITLE: Integration of Text, Image, and Chemistry Databases on GUI Clients Designed for Drug Discovery and Development

PRINCIPAL INVESTIGATOR: Patrick McGreevy, Ph.D.

CONTRACTING ORGANIZATION: Dataaspects Corporation
Los Gatos, California 95032

REPORT DATE: June 1997

TYPE OF REPORT: Final - Phase I

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1997	Final-Phase I (8 Nov 96 - 7 May 97)

4. TITLE AND SUBTITLE
Integration of Text, Image, and Chemistry Databases on GUI Clients Designed for Drug Discovery and Development

5. FUNDING NUMBERS
DAMD17-97-C-7018

6. AUTHOR(S)
Patrick McGreevy, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dataaspects Corporation
Los Gatos, California 95032

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution authorized to U.S. Government agencies only (specific authority). Other requests for this document shall be referred to Commander, U.S. Army Medical Research and Materiel Command, ATTN: MCMR-RMI-S, Fort Detrick, Frederick, Maryland 21702-5012.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The purpose of this effort was to create a prototype Chemical Information System that integrates chemical, biological and inventory databases into a Graphical User Interface (GUI) for drug discovery. TRIMtools, a 4th generation C like language from Trifox Inc, were used to create the GUI because of their low cost, high performance, platform independence and library functions for database management across networks. Thin and thick client GUIs were written for Windows NT/95/3.1 and a Java Client written for all major platforms including MacOS. These clients manage simultaneous connections to a Daylight chemical database on a Sun Solaris, an Rdb database on a Vaxstation and an Image database on a jukebox. Data from each server are merged within the GUI window or displayed by helper applications. Both 2D and 3D chemical structures, Thor Datatrees, and Merlin similarity searches from the Daylight CGI are displayed in a Netscape browser and a RasMol window. Images of Chemical Data Sheets stored as tiff files and reference monographs stored as PDF files were displayed in graphical viewers. The need to code printer specific reports was eliminated by writing text and images into dynamic HTML that is printed from the client's browser. These features were used to format custom screens to enable biologists, chemists and supporting staff to manage high throughput analyses typical of the pharmaceutical industry.

14. SUBJECT TERMS
drug development, chemical information, chemical disposal, database management, client server, GUI, networks, web

15. NUMBER OF PAGES
12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Limited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

PBMcA Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Patrick B McAreavy 6 June 97

PI - Signature Date

Table of Contents

Front Cover	1
Standard form 298	2
Foreword	3
Table of Contents	4
Introduction	5
Task 1. Move WR CIS from Oracle Rdb to Informix Online	6
Task 2. Program Image Windows into the TRIM GUI Client	6
Task 3. Interface TRIM GUI Client and the Optical Image Repository	7
Task 4. Write TRIMtool functions for HTTP Exchanges with Daylight CGI	8
Task 5. Upgrade the WR CIS to support fat & thin TRIM GUI Clients	9
Task 6. Develop an HTML Document Writer as a Print Utility for GUI Clients	9
Task 7. Port GUI Client to Macintosh	10
Task 8. Prepare reports & WR demonstration	10
Conclusions	10
Meeting Abstract	12
Personnel	12

INTRODUCTION

Data management in the pharmaceutical, agrochemical and high technology industries is complex because diverse types of data from chemical, textual and image databases which often run on different platforms must be merged into information that supports research and development. The automated integration of these databases into useful information on the biological activity of chemicals along with relevant reference materials and inventory transactions is a difficult task that few companies have achieved. As a consequence, database integration is typically a manual process or semi manual process. The purpose of this initiative was to explore techniques to integrate information from chemical, biological and image databases into a client Graphical User Interface (GUI) designed for drug discovery and development.

Research and development on the new GUI was conducted as a collaborative effort between DataAspects Inc., Trifox Inc (www.trifox.com), Daylight Chemical Systems (www.daylight.com), and the Department of Chemical Information at the Walter Reed Army Institute of Research (http://wrair-www.army.mil). A prototype of the CIS was created with TRIMtools from Trifox which consist of default scripts to display client windows and a function library for connecting to databases, managing data, and distributing it over networks including the Internet. The CIS prototype was built on multiple databases residing on different servers: (1) an RDB database containing 3.4 million records on a Vaxstation in Maryland, (2) an Informix database containing the same 3.4 million records on a Sun Solaris in Maryland, (3) a Daylight database that manages 300,000 chemical structures on the Sun Solaris in Maryland, and (4) an image database containing 90,000 tiff files on an optical jukebox controlled by a Vaxstation in Virginia. The actual GUI applications were created in California on a Windows NT client using simultaneous Internet connections to these servers.

The central application in the GUI prototype is a single client window called Start that connects to three databases simultaneously and provides the user with access to all other applications via push buttons and drop down menus. The integration of imaged data was achieved through the use of third party viewers. For example, 2D and 3D chemical structures are displayed by sending chemical identifiers (SMILES) to the Daylight CGI which returns an image of the structure for display in the Netscape browser. Likewise, images of chemical Data Sheets and reference monographs are retrieved from the jukebox and displayed in third party image viewers. In addition to viewing images on the client's display, these helpers also print the images on the client's local printer. This technique works so well that all reports are now formatted into HTML documents and printed via the Netscape browser.

TRIMtools can be configured to run as a fat client with all of programs installed on the client computer, or a thin client with only three files occupying less than 100 kb of disk space. All development of the CIS prototype was done on a Windows NT computer configured as a thick client. The CIS prototype has been run under the thin client configuration installed on three operating systems (WinNT, Win95, Win3.1). As part of

Phase I, a JAVA client was written which will run on all major platforms including Macintosh.

A detailed description of the progress achieved under each of the tasks proposed in the Phase I protocol follows.

TASK 1. Move WR CIS from Oracle Rdb to Informix Online

This task was completed in collaboration with John Notsch, Department of Chemical Information. Using TRIMtools from Trifox Inc, some 3.4 million records of data were exported from Rdb and imported into Informix. The Informix database is accessible from the network and will be used to develop new GUI applications. The Department of Chemical Information plans to put the new Informix database on line in late 1997.

TASK 2. Program Image Windows into the TRIM GUI Client

TRIMtools has a function called winexec() that calls an executable program on the client computer and passes it a parameter. To display an image, winexec() invokes an image viewer and passes it the appropriate image file which is then projected in a separate window. To complete this task, DataAspects utilized third party imaging software that is typically found on client computers or is available free of charge. In its current configuration, the CIS must use four different image viewers to display three types of image files, compressed tiff files of Chemical Data Sheets, PDF files of online publications, gif files of 2D chemical structures and a chemical MIME type of 3D structures. Since the projection of chemical images involves special software, it is described separately under Task 4.

Tiff files are displayed using imaging software written by Wang that is available free from Microsoft Corporation and included as part of its Windows NT and 95 operating systems. To view a Data Sheet, the Bottle Number is read from the client's display and used to search an index table for the address of the corresponding tiff file on an optical jukebox. This address is passed to an ftp like TRIM function called file_copy() that gets the tiff file from the jukebox, transmits it over the network, and writes it to the CIS cache directory on the clients hard drive. Finally, winexec() calls the Wang imager and passes it the address of the cached tiff file which is displayed in a separate 'Wang Window'. This scenario is discussed in more detail under Task 3.

Over the years the Division of Experimental Therapeutics has produced three important monographs on the chemotherapy of malaria, chemotherapy of leishmaniasis, and anti-radiation drugs. These documents archive considerable biological and chemical information on the chemicals in the WR database. There are limited copies of these monographs and they are not readily available to the scientific community. As an adjunct to Phase I, DataAspects has studied the feasibility of publishing these documents online in a format that can be searched and displayed from CIS applications. This task is

particularly difficult because the documents contain images of chemical structures surrounded by text and wide tables that span multiple pages. Adobe Acrobat PDF files were able to store text, structures and wide tables in files that could be transmitted over a network in a reasonable time and displayed, navigated and searched in Adobe Acrobat Reader running on the client machine. It is important to note that the PDF file is an industry standard for online publishing and that the free Acrobat runs on all major platforms including Windows NT/95/3.1 and Macintosh. Like the tiff files described above, the user can download specific PDF files from a distant server over a network to the local cache directory and view them in the Acrobat Reader. This scenario is all controlled from within the CIS applications. Having shown that online publishing of the monographs is feasible and that online documents can be integrated into the CIS, a plan to convert them from paper and magnetic tape to PDF files will be proposed as a Phase II task.

A second pilot study on imaging was also conducted to determine the feasibility of integrating mass spectra data on +/-30,000 WR compounds into the CIS. Each mass spectrum is stored as x/y coordinates with associated header data. As part of Phase II, DataAspects will propose to store these data in a database table that can be searched on Bottle Number from a CIS application. The selected mass spectrum data will be passed to a graphics program and the resulting graph converted to a gif or PDF image that can be downloaded to the client's cache directory and displayed in the browser.

A final note on imaging is worthy of note. As the user works in the new CIS, image files accumulate in the CIS cache directory on the client computer. For the sake of speed, the CIS searches this client directory for a requested file before getting the file over the network from a distant server. While this technique increases performance, it also compromises security as files containing proprietary data are distributed on client computers throughout the corporation. To protect private information, all files in the cache directory are deleted when the user exits the CIS.

TASK 3. Interface TRIM GUI Client and the Optical Image Repository

Chemical Data Sheets are stored in individual tiff files on a jukebox controlled by a server Vaxstation in Virginia. There are 500,000+ Data Sheets and 90,000 of them have been scanned into individual tiff files while the remainder are being scanned. The first problem in retrieving a particular tiff file is to locate its physical position on the jukebox. To resolve this issue, an import program was written that reads the directories of the jukebox, parses each filename into 'Bottle Number' and 'Tiff address' fields, and inserts them into an index table in the CIS relational database. By indexing on 'BN' this table serves as an index to look up a 'BN' and retrieve the physical path to the corresponding Tiff file on the Herner server.

Once the path is known, it must be copied from the jukebox to the CIS cache directory on the client computer. To achieve this file transfer, a TRIMtools function called

file_copy() was modified to get files over networks by communicating with a server daemon running on the Server Vaxstation.

This is how it works. A user displays CIS data and pushes the 'Data Sheet' action button. An action button has been added to GUI applications that downloads a tiff file from the Herner server and displays it with client imaging software. First the BN from the client's display is sent to the CIS server in a SQL select statement and the path to the appropriate Data Sheet is retrieved. This address is used as a parameter in the file_copy() function to download the tiff image of the Data Sheet to the client cache directory. Finally, the Wang imaging software is called and the image is displayed in a separate image window. The user can print this image to the clients computer via the Wang printing utility.

TASK 4. Write TRIMtool Functions for HTTP Exchanges with Daylight CGI

The integration of textual data from the CIS relational database with chemical data from the Daylight Chemical database is mediated through Daylight's CGI (Common Gateway Interface). This interface offers simple access to Daylight's Thor database and Merlin searching utilities.

A user can display structures of the current chemicals on the client display by pushing the 'Structure' action button which calls a TRIM function to assemble the SMILES into a list in memory. Duplicate SMILES are removed and the remaining list of distinct SMILES is sent to a second function that converts them into hexadecimal code. The hex encoded SMILES are concatenated with the URL pointing to the Daylight CGI and formatted into a table within an HTML document. This HTML document is opened in the client's Netscape browser using a simple call from the CIS application. As the browser reads the HTML table it automatically sends the hex encoded SMILES to the Daylight CGI which returns a depiction of the structures as gif files. Finally, the gif files are displayed in the browser window inside the HTML table. On average it takes 3 seconds to open the browser and display a structure on a client in California communicating at 14.4 KBPS with the Daylight CGI on the Sun Workstation in Maryland.

A variation of the above scenario was implemented to display Thor Data Trees and 3D structures from the HTML table of structural gif files. Each of these gif files is superimposed with three client side image maps that link to the Daylight CGI. The user can click the top image map to display the Thor Data Tree, the middle image map for a 3D display without hydrogen, or the bottom image for a 3D display with hydrogen. To produce 3D structures the browser sends the hex encoded SMILES of the selected 2D structure to the Daylight CGI which returns the 3D coordinates of the atoms in a stream of MIME type 'chemical/x-cex-3D'. This stream is passed to a chemical viewer called RasMol, (free shareware from http://www.umass.edu/microbio/rasmol/index.html), where it is displayed in 3D and rotated by the user with the mouse.

Using the technology described above, the following Daylight CGI scripts have been written and linked to the Start application: Merlin similarity search, Merlin

superstructure search, and Thor Data Trees (TDT) . In addition, a script that links to chemical databases other than those maintained by WR has also been added. By clicking the appropriate action button, Start calls custom these custom scripts in the Daylight CGI and passes the appropriate hex encoded SMILES as a parameter. The CGI returns HTML documents with the appropriate information.

Finally, GRINS, the Daylight molecular editor, has been integrated into the CIS and is called from an action button in the Start and Accession applications. By using GRINS, the user can construct a structure on the screen and convert it into a unique SMILES that can be used to query the CIS database for biology and inventory information. SMILES queries are new to the CIS and are implemented through a database table relating SMILES with WR Numbers. Although it is invisible to the user, a SMILES query is really translated into a WR Number query using this index table.

TASK 5. Upgrade the WR CIS to Support Fat & Thin TRIM GUI Clients

The intent of Phase I was to determine if the new GUI could be used to integrate chemical, biological and inventory data into organized screens across networks and not to convert the 50+ text mode applications currently used at WR into GUI mode. However, five of the text mode applications have been converted to GUI mode during the creation of the CIS prototype. A description of this prototype follows.

The opening GUI application is called Start and it replaces the text base application called TRIMmenu. Start opens with a Connect Window where the user enters the username/password authentication. The Connect Window uses the username/password and the IP addresses stored in an ASCII file called WRCIS.INI to construct connect strings used to log into the CIS servers. If login is successful, the Connect Window calls the Enter Window which is central to the new CIS because it contains menu items, radio buttons, check boxes and drop down lists that lead to all utilities in the entire WR CIS. For example, the user can call the Hxship Window to see the history of shipping for a particular Bottle Number or the Request Window to request the shipment of a chemical sample. The user can also call other applications such as Accession to register a new chemical sample into the WR repository, Merck to cross index Merck Numbers with WR Numbers, and Track to follow a compound through the ordering/shipping system.

Having shown that GUI TRIMtools provide solutions to all of the chemical information needs of the Army Drug Discovery team, the conversion of the remaining 45+ text mode applications in the GUI mode will be proposed as a SBIR Phase II project.

TASK 6. Develop an HTML Document Writer as a Print Utility for GUI Clients

A function has been written that will convert the contents of a TRIM list into HTML. By calling this function, data is read from the GUI screen and reformatted into an HTML document that is displayed in the Netscape browser. Using the Netscape print utility, the document will be sent to the client's local printer.

An example of this HTML utility, is the creation of a WR Chemical Data Sheet from the Accession application. The WR Data Sheet has been redesigned and includes the structure of the molecule and a box for free text for user comments. After selecting the 'Create DS' action button, the current SMILE is sent to the Daylight CGI which returns the gif image of the structure. This gif image and the textual data displayed in the Accession application are formatted dynamically into an HTML document and displayed in the Netscape browser. Finally, this HTML Data Sheet can be printed locally from the browser.

TASK 7. Port GUI Client to Macintosh

DataAspects has written thin client software using Java that is supported on all major platforms including the Macintosh. At present, the Java client will run the WR CIS on the Macintosh, but some of the widgets like 3D have not been implemented. It is anticipated that the JAVA implementation will be completed by the end of June, 97. The Java implementation is particularly important to the CIS manager as his endless task to maintain software on client computers will be minimized if not eliminated.

TASK 8. Prepare Reports & WR Demonstration

Interim reports have been submitted for months 1-5 of Phase I. An early prototype of the new CIS was demonstrated at the Daylight User's Group meeting in late February and the final prototype of was demonstrated at the Division of Experimental Therapeutics in early May, 1997. It is not an overstatement to say that the new CIS was met with enthusiasm. One of the participants at the User's group stated that Pat McGreevy would be a hero if he were to demonstrate the CIS at his company and this contact could be DataAspects' first customer.

CONCLUSIONS

The GUI implementation of TRIMtools for the creation of the CIS prototype demonstrates capabilities that far exceed the objectives listed in the Phase I SBIR proposal. Perhaps the most significant finding during Phase I was the ease that TRIMtools integrated with commercial software from third party vendors to create powerful applications that organize diverse data into information on a single screen. The following list of features built into the WR CIS prototype clearly demonstrates its power and flexibility.

1. TRIMtools Server software runs on 13 operating systems: AIX, HP/UX, MVS, OS/2, OSF/1, (DEC UNIX), OpenVMS, SCO, Solaris, SunOS, Windows, Windows 95, Windows NT.

2. Thin and Thick Client software provides a graphical user interface that runs on 13 platforms: HP/UX, Linux, MVS, MacOS, S/2, OSF/1 (DEC UNIX), OpenVMS, SCO, Solaris, SunOS, Windows, Windows 95, Windows NT . Thin Client runs on very small PCs like 386-based Windows 3.1 systems with 8Mb of RAM to high end PCs running Win95/NT and Work Stations. The Java GUI which will be deployed in late June will run on all of the platforms and the MacOs.

3. Default C based GUI applications are created quickly using 4GL features. Customization is facilitated by access to application code and the availability of a function library focused on database management and the manipulation of data lists in memory.

4. Connections to 13 commercial relational databases are supported: ADABAS C, ADABAS D, DB2, Informix, Ingres, ODBC, Oracle, Rdb, SDMS (ISAM), SQL Server, Sybase, c-tree (ISAM) and dBase.

5. Six connections to different servers can be opened simultaneously to access textual, chemical and image databases stored on different computers.

6. Tiff files on an optical jukebox can be accessed using an ftp like function and displayed in image windows on the client computers with a helper programs. Likewise, PDF files can be retrieved across networks and displayed on client computers using the Adobe Acrobat Reader.

7. Chemical data can be accessed by sending SMILES from the GUI applications to scripts in the Daylight Common Gateway Interface. The results are displayed as HTML documents in the client browser. Current utilities include the display of 2D structures, Thor Data Trees, Merlin similarity searches and GRINS. RasMol is used as a helper application to display structures in 3D.

8. By formatting text and image data into dynamic HTML documents, printing responsibilities are passed from the GUI applications to the client's browser eliminating the need to code reports for platform specific clients and local printers.

The GUI implementation of the WR CIS is a low-cost, high-performance system accessible from inexpensive terminals and PCs connected to an Intranet or the Internet. There is no doubt that the integration of biology, chemistry and inventory data into simple screens will increase productivity by providing relevant information as it is generated to each member of the research team from the technical staff to the Director.

MEETING ABSTRACTS

Pat McGreevy presented 'The Walter Reed Chemical Information System' at the User's Group Meeting hosted by Daylight Chemical Information Systems in Laguna Beach

California on 15-28 February, 1997. The abstract for this meeting can be found at the following URL: http://www.daylight.com/meetings/mug97/agenda97/McGreevy/

PERSONNEL

The following personnel conducted this study and received pay for their efforts:

Patrick B. McGreevy
Paul Dembry
Niklas Back

DEPARTMENT OF THE ARMY
US ARMY MEDICAL RESEARCH AND MATERIEL COMMAND
504 SCOTT STREET
FORT DETRICK, MARYLAND 21702-5012

REPLY TO
ATTENTION OF:

MCMR-RMI-S (70-1y) 4 Dec 02

MEMORANDUM FOR Administrator, Defense Technical Information Center (DTIC-OCA), 8725 John J. Kingman Road, Fort Belvoir, VA 22060-6218

SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has reexamined the need for the limitation assigned to technical reports written for this Command. Request the limited distribution statement for the enclosed accession numbers be changed to "Approved for public release; distribution unlimited." These reports should be released to the National Technical Information Service.

2. Point of contact for this request is Ms. Kristin Morrow at DSN 343-7327 or by e-mail at Kristin.Morrow@det.amedd.army.mil.

FOR THE COMMANDER:

Encl

PHYLIS M. RINEHART
Deputy Chief of Staff for
Information Management

ADB218773
ADB223531
ADB230017
ADB223528
ADB231930
ADB226038
ADB224296
ADB228898
ADB216077
ADB218568
ADB216713
ADB216627
ADB215717
ADB218709
ADB216942
ADB216071
ADB215736
ADB216715
ADB215485
ADB215487
ADB220304
ADB215719
ADB216072
ADB222892
ADB215914
ADB222994
ADB216066
ADB217309
ADB216726
ADB216947
ADB227451
ADB229334
ADB228982
ADB227216
ADB224877
ADB224876
ADB227768
ADB228161
ADB229442
ADB230946
ADB230047
ADB225895
ADB229467
ADB224342
ADB230950
ADB227185
ADB231856
ADB229914
ADB229497
ADB230947
ADB282209
ADB270846
ADB282266
ADB262442
ADB256670